LA-UR-19-28986
Approved for public release; distribution is unlimited.

Title: Nuclear Data Dimension Reduction

Author(s): Francom, Devin Craig
Vander Wiel, Scott Alan
Weaver, Brian Phillip

Intended for: Report

Issued: 2019-09-06

Disclaimer:
Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Nuclear Data Dimension Reduction

Devin Francom, Scott Vander Wiel, Brian Weaver
CCS-6, Statistical Sciences
Los Alamos National Laboratory

1 Introduction

Our nuclear data consist of 30 energy group cross sections as well as nubar and PFNS. They are sampled from a Gaussian distribution with correlations shown in Figure 1. A few of these samples are shown in Figure 2. The purpose of this section is to describe how we can reduce the dimension of the nuclear data for purposes of uncertainty quantification. Sections 2 and 3 describe why dimension reduction is useful, as well as our approach to dimension reduction.

When using these nuclear data as inputs to the PARTISN code, some of the resulting k_{eff} values are unrealistic. For instance, when PARTISN is simulating the Jezebel critical assembly, some of the nuclear data yield k_{eff} far from one. We would like to weight the nuclear data combinations that yield more realistic output higher, while downweighting those that result in unrealistic output. Our approach to this is described in Section 4. Throughout the other sections, we will merely refer to a prior distribution of nuclear data, which is not constrained by the Jezebel experiment, and a posterior distribution of nuclear data, which is conditioned on satisfactory performance for the Jezebel experiment. For instance, the correlation matrix for the reweighted (posterior) nuclear data is shown in Figure 3, along with the difference from the prior correlation matrix.

Finally, we will describe dimension reduction results in Section 5.

2 Dimension Reduction in Regression Settings

In regression settings, high dimensional inputs can be debilitating. Consider the case where we have a p-dimensional set of inputs, $\mathbf{x} = (x_1, \ldots, x_p)$ that produce output $y = f(\mathbf{x})$. In order to explore the input-output relationship (i.e., to explore f), we might evaluate y for a collection (say N) of different combinations of $\mathbf{x}$. The difficulty with exploring this relationship comes from the curse of dimensionality (Donoho et al. (2000)), which is that as p increases linearly, N must increase exponentially in order to explore the larger space. Dimension reduction is typically used in an effort to combat the curse of dimensionality.

There are numerous flavors of dimensions reduction, all with the same goal of representing high-dimensional data in the most informative lower dimensional space. More

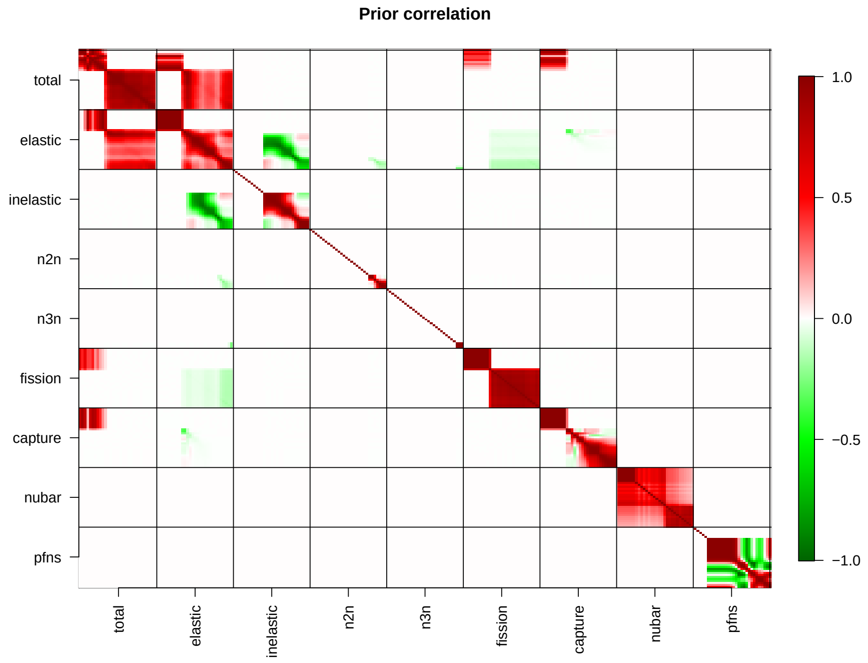

Figure 1: Visualization of the nuclear data correlation matrix.

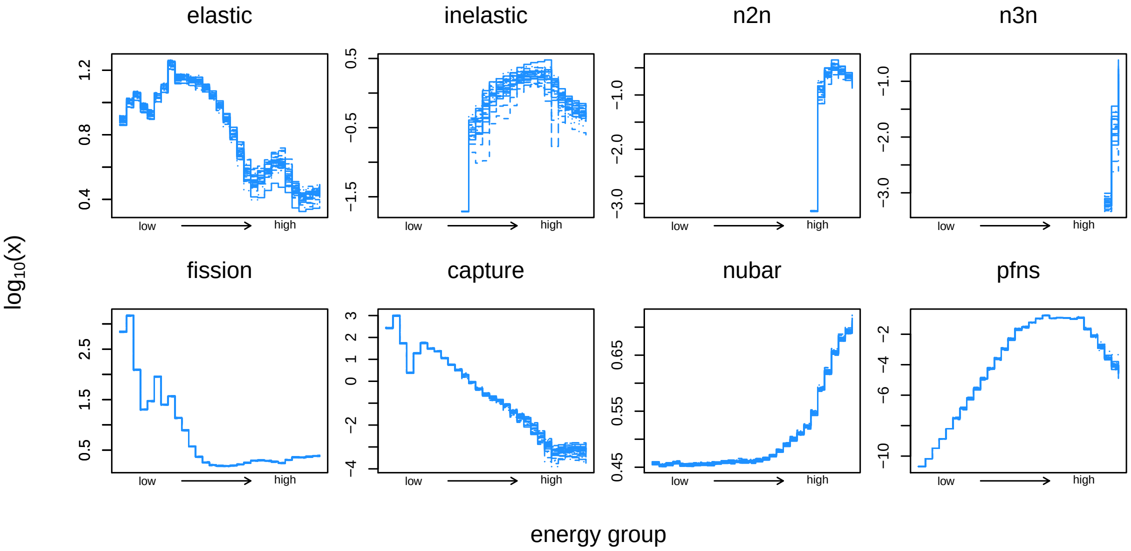

Figure 2: 30 samples of the nuclear data ($\log_{10}$ scale).

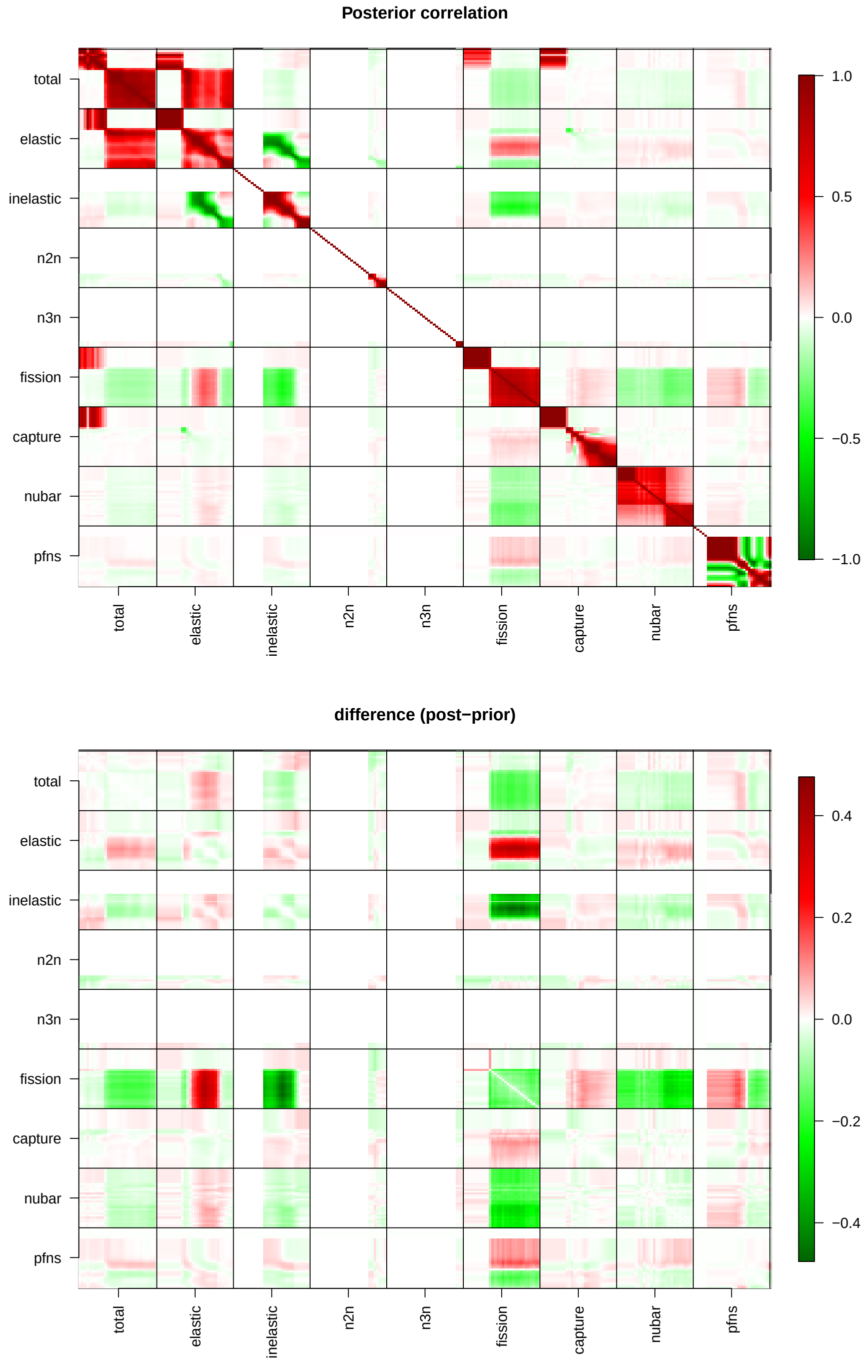

Figure 3: Top: Correlation matrix of the nuclear data after reweighting according to the Jezebel experiment. Bottom: difference (posterior-prior) between the correlation matrices.

specifically, given the $\mathbf{x}$-y relationship above, we are interested in finding a reduced dimension set of inputs $\mathbf{z} = (z_1, \ldots, z_d)$ where $d < p$ and $y \approx g(\mathbf{z})$. In other words, we want to find the reduced dimension input space that still explains all (or the majority) of the variation in y, though through a different functional relationship. If we can discover such a reduced dimensional input, then the function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ will be easier to explore than $f : \mathbb{R}^p \rightarrow \mathbb{R}$.

A natural approach to determining $\mathbf{z}$ is to consider linear subspaces, which have the form $\mathbf{z} = (\mathbf{b}_1'\mathbf{x}, \ldots, \mathbf{b}_d'\mathbf{x}) = \mathbf{B}'\mathbf{x}$. The vectors $\mathbf{b}_1, \ldots, \mathbf{b}_d$, which are typically orthogonal, are often called the active directions. The space spanned by the active directions is of primary interest, so the scale of each $\mathbf{b}$ vector is uninteresting. There are numerous approaches to finding the active directions. Below we explore a few of these approaches, namely principal components, sliced inverse regression, and active subspaces.

2.1 Principal Component Approach

The principal component approach to this problem is to set $\mathbf{b}_1, \ldots, \mathbf{b}_d$ equal to the dominant eigenvectors of the matrix $Cov(\mathbf{x})$, where $Cov(\cdot)$ denotes covariance. Then $\mathbf{b}_1$ can be interpreted as the direction corresponding to largest variation in $\mathbf{x}$, $\mathbf{b}_2$ as the direction with largest variation orthogonal to $\mathbf{b}_1$, and so on. Since the eigenvalues reveal the amount of variance of $\mathbf{x}$ explained by the different eigenvectors, d can be chosen according to some cutoff of percent of variance explained.

Because the principal components are the dominant modes of variation in $\mathbf{x}$ without regard to y, the principal component approach is not specific to regression settings. While general and widely used for dimension reduction in the regression setting, other approaches can sometimes be more appropriate because of their exploration of variation in both $\mathbf{x}$ and y.

2.2 Sliced Inverse Regression (SIR)

Sliced inverse regression methods, introduced in Li (1991), seek to find $\mathbf{b}_1, \ldots, \mathbf{b}_d$ such that $p(y|\mathbf{x}) \approx p(y|\mathbf{b}_1'\mathbf{x}, \ldots, \mathbf{b}_d'\mathbf{x})$. That is, SIR seeks the directions of $\mathbf{x}$ that explain the dominant variation in y.

To find $\mathbf{b}_1, \ldots, \mathbf{b}_d$, consider reversing the conditioning, to think about $p(\mathbf{x}|y)$. More specifically, consider using $E(\mathbf{x}|y)$, called the inverse regression curve, as a summary of the relationship between $\mathbf{x}$ and y. To explain the intuition behind the inverse regression curve, consider an example when $p = 3$. For $y = y_0$, the multivariate distribution of (x_1, x_2, x_3) may be constrained in x_1 and x_2 but not in x_3. That is to say that y_0 can only be obtained when x_1 and x_2 are close to certain values, regardless of the value of x_3. Under a few conditions, this relationship is captured by $E(\mathbf{x}|y)$. Let $\tilde{\mathbf{x}} = \mathbf{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$, where $\boldsymbol{\mu} = E(\mathbf{x})$ and $\mathbf{\Sigma} = Cov(\mathbf{x})$. Then the dominant eigenvectors of $Cov(E(\tilde{\mathbf{x}}|y))$, which we call $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_d$, are the dominant modes of variation of the inverse regression curve. Thus, we could use $\mathbf{\Sigma}^{-1/2}\boldsymbol{\eta}_1, \ldots, \mathbf{\Sigma}^{-1/2}\boldsymbol{\eta}_d$ as $\mathbf{b}_1, \ldots, \mathbf{b}_d$.

In practice, the inverse regression curve is approximated using a slicing technique, and hence the complete method is called “sliced inverse regression.” The slicing is performed by breaking the range of y into distinct bins and, for each bin, considering the sample mean of the $\tilde{\mathbf{x}}$ vectors that result in a y in that bin. Specifically, for bin h, we calculate the mean of the corresponding samples, $\bar{\mathbf{x}}_h$, and the number of samples that fall in that bin, n_h. Then we estimate $Cov(E(\tilde{\mathbf{x}}|y)) \approx \frac{1}{N}\sum_{h=1}^{H} n_h \bar{\mathbf{x}}_h \bar{\mathbf{x}}_h'$. Figure 4 demonstrates what a few $\bar{\mathbf{x}}_h$ vectors look like for the nuclear data. Different colors indicate different bins (different values of h). This indicates, for instance, that the high energies of the fission cross section tend to explain differences in y.

2.3 Active Subspaces

The active subspace approach to this problem (Russi (2010); Constantine et al. (2014)) uses the gradient of f, denoted $\nabla_x f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_p}\right)'$ to understand the relationship between $\mathbf{x}$ and y. To find the dominant modes of variation in the gradient, we use the eigen decomposition of $E[(\nabla_x f)(\nabla_x f)']$, which is the (typically approximated) uncentered covariance of the gradient vector. In cases where the gradient can be sampled, this is approximated with $E[(\nabla_x f)(\nabla_x f)'] \approx \frac{1}{N}\sum_{i=1}^{N}(\nabla_x f_i)(\nabla_x f_i)'$. In cases where the gradient cannot be sampled explicitly, it can be approximated via further sampling (Constantine et al. (2015)).

When the dimension reduction results in $d = 1$ (i.e., $z = \mathbf{b}'\mathbf{x}$), and when the relationship between z and y is linear (i.e., $g(z) = \alpha_0 + \alpha_1 z$), then the problem simplifies to linear regression where we can find $\mathbf{b}$ by least squares.

3 Gaussian Regularized SIR

In cases where we think that an active direction $\mathbf{b}$ should be smooth with respect to an indexing variable, regularization becomes necessary. Bernard-Michel et al. (2009) introduce a method for regularizing SIR particularly for the case when $\mathbf{\Sigma}$ is not invertible. The idea is to regularize by effectively assuming a Gaussian prior for $\mathbf{b}$.

In Section 2.2, we described how the SIR subspace of interest is obtained by taking the first d transformed eigenvectors of $Cov(E(\tilde{\mathbf{x}}|y))$. The same subspace is obtained by taking the first d eigenvectors of $\mathbf{\Sigma}^{-1}Cov(E(\mathbf{x}|y))$. Let $\mathbf{\Gamma} := Cov(E(\mathbf{x}|y))$. Assuming the Gaussian prior for each vector $\mathbf{b}_i$ is of the form $\mathbf{b}_i \sim N(\mathbf{0}, c\mathbf{\Omega})$, where c is a constant, then, under the inverse regression model introduced in Cook (2007), the posterior maximum for $\mathbf{b}_i$ is the i^{th} largest eigenvector of $(\mathbf{\Omega}\mathbf{\Sigma} + \mathbf{I})^{-1}\mathbf{\Omega}\mathbf{\Gamma}$. Bernard-Michel et al. (2009) demonstrate that three previous approaches to SIR fit into this formulation: (1) traditional SIR is obtained using $\mathbf{\Omega}_0 = (\tau\mathbf{\Sigma})^{-1}$; (2) a ridge regularized version obtained using $\mathbf{\Omega}_1 = \tau^{-1}\mathbf{I}$; and (3) SIR on principal components of $\mathbf{x}$ is obtained using $\mathbf{\Omega}_2 = \tau^{-1}\sum_{j=1}^{q}\mathbf{u}_j\mathbf{u}_j'/\delta_j$, where $\mathbf{u}$ is an eigenvector, δ is an eigenvalue, and $d \leq q \leq p$. In all of these, $\tau > 0$ is an unknown parameter. Bernard-Michel et al. (2009) also introduce three new SIR regularizations: (1) Tikhonov regularization is introduced with $\mathbf{\Omega}_3 = \tau^{-1}\mathbf{\Sigma}$; (2) ridge regularization on the PCA projected predictors is introduced with $\mathbf{\Omega}_4 = \tau^{-1}\sum_{j=1}^{q}\mathbf{u}_j\mathbf{u}_j'$; and (3) Tikhonov regularization on the PCA projected predictors is introduced with $\mathbf{\Omega}_5 = \tau^{-1}\sum_{j=1}^{q}\delta_j\mathbf{u}_j\mathbf{u}_j'$.

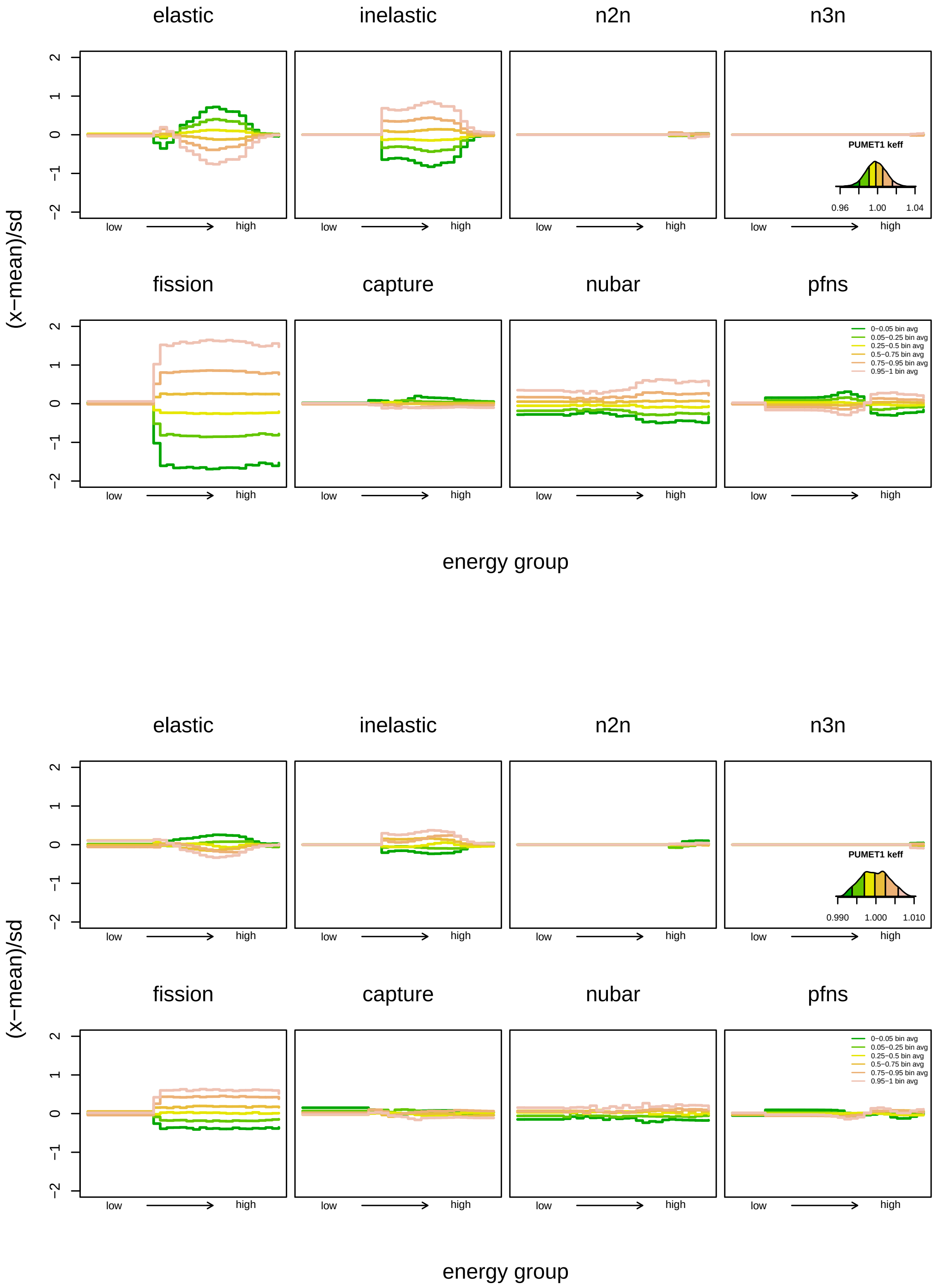

Figure 4: Each line is the average of cross section samples that produce y values that fall into a particular bin. Hence, the darkest green line corresponds to the smallest y values. The top group of plots show the nuclear data with prior weights, whereas the bottom shows the reweighted version with the same axes, demonstrating the decrease in variation.

3.1 Smoothness

Consider the case where $\mathbf{b}$ is expected to be smooth with respect to index t. Thus the i^{th} element of $\mathbf{b}$ is $b_i = b(t_i)$. To impose smoothness we can regularize the second derivative of $b(t)$. That is, we can penalize $b''(t) := \frac{d^2b(t)}{dt^2}$ being different from zero, with the prior $b''(t) \sim N(0, \kappa^{-1})$. In the discretized case, we can approximate $b''(t)$ with $\mathbf{Kb}$ where $\mathbf{K}$ is the $(T-2) \times T$ second differencing matrix where T is the length of $\mathbf{b}$. Hence, the first row of $\mathbf{K}$ is $(-1, 2, -1, 0, \ldots, 0)$, the second row of $\mathbf{K}$ is $(0, -1, 2, -1, 0, \ldots, 0)$ and the last row of $\mathbf{K}$ is $(0, \ldots, 0, -1, 2, -1)$. Then $\mathbf{Kb} \sim N(\mathbf{0}, \kappa^{-1}\mathbf{I})$, which implies $\mathbf{b} \sim N(\mathbf{0}, (\kappa\mathbf{K}'\mathbf{K})^{-})$, where $\mathbf{A}^{-}$ is the generalized inverse of $\mathbf{A}$. However, we can also combine the smoothness prior with the other priors discussed above, as they represent two independent sources of prior information. We propose three new SIR regularizations that include penalties for lack of smoothness, parameterized in terms of precision rather than covariance: (1) Smooth SIR with $\mathbf{\Omega}_6^{-1} = \tau\mathbf{\Sigma} + \kappa\mathbf{K}'\mathbf{K}$; (2) smooth ridge with $\mathbf{\Omega}_7^{-1} = \tau\mathbf{I} + \kappa\mathbf{K}'\mathbf{K}$; and (3) smooth Tikhonov with $\mathbf{\Omega}_8^{-1} = \tau\mathbf{\Sigma}^{-1} + \kappa\mathbf{K}'\mathbf{K}$.

We want to allow for the possibility that $\mathbf{\Sigma}$ is singular and therefore $\mathbf{\Omega}^{-1}$, above, is not well-defined. A singular $\mathbf{\Sigma}$ can arise because of the functional form of $\mathbf{x}$. Using the Woodbury matrix identity on $\mathbf{\Omega}_8$, for example,

$$\begin{aligned}
\mathbf{\Omega}_8 &= \left(\tau\mathbf{\Sigma}^{-1} + \kappa\mathbf{K}'\mathbf{K}\right)^{-1} \\
&= \left[\left(\tau^{-1}\mathbf{\Sigma}\right) - \left(\tau^{-1}\mathbf{\Sigma}\right)\mathbf{K}'\left[\kappa^{-1}I + \mathbf{K}\left(\tau^{-1}\mathbf{\Sigma}\right)\mathbf{K}'\right]^{-1}\mathbf{K}\left(\tau^{-1}\mathbf{\Sigma}\right)\right] \\
&= \tau^{-1}\mathbf{\Sigma}\left[\mathbf{I} - \kappa\mathbf{K}\left(\tau\mathbf{I} + \mathbf{K\Sigma K}'\right)^{-1}\mathbf{K\Sigma}\right].
\end{aligned}$$

This latter form of $\mathbf{\Omega}$ is computable even if $\mathbf{\Sigma}$ is not invertible. Further, singular $\mathbf{\Sigma}$ implies singular $\mathbf{\Omega}$ but the GRSIR method can proceed because $(\mathbf{\Omega\Sigma} + \mathbf{I})^{-1}\mathbf{\Omega\Gamma}$ remains well-defined, as does its eigen-decomposition.

In the case where $\mathbf{b}$ is a collection of r functions indexed by the same variable (i.e., $\mathbf{b} = [b_1(t_1), \ldots, b_1(t_T), b_2(t_1), \ldots, b_2(t_T), \ldots, b_r(t_T), \ldots, b_r(t_T)]'$), we apply the smoothness prior to each function independently. Hence, we make the matrix $\mathbf{K}_r = \mathbf{I}_r \otimes \mathbf{K}$, where $\otimes$ is the Kronecker product, $\mathbf{I}_r$ is the $r \times r$ identity matrix, and $\mathbf{K}$ is the same $(T-2) \times T$ matrix defined above. We then replace $\mathbf{K}'\mathbf{K}$ in $\mathbf{\Omega}_6$, $\mathbf{\Omega}_7$ and $\mathbf{\Omega}_8$ with $\mathbf{K}_r'\mathbf{K}_r = \mathbf{I}_r \otimes (\mathbf{K}'\mathbf{K})$.

3.2 Parameter Choices

For a given choice of $\mathbf{\Omega}$, the parameter τ needs to be estimated in all cases as does the parameter κ in the smoothing cases. We do this by cross-validation. However, cross-validation is not straight forward, since the relationship, $g(\cdot)$, between the active variables, $\mathbf{z}$, and the response, y, is unknown. With a large number of function evaluations relative to the dimension of $\mathbf{z}$, estimating $g(\cdot)$ with nearest neighbors is attractive for simplicity. Using a more complex model would likely require much more training time for each parameter setting in the cross validation. Further, a more complex model would likely have parameters of its own to tune, possibly confounding the cross validation.

We use 10-fold cross validation to choose τ and κ . For reasons discussed in the Introduction and further explained in Section 4.1, we have a weighted sample of $\mathbf{x}$ vectors, $\mathbf{x}_1, \ldots, \mathbf{x}_M$ with corresponding weights $w_1, \ldots, w_M$. We break the M samples into 10 folds requiring each to have similar weight (similar effective sample size, a quantity discussed below). In both the training and test set, we resample the $\mathbf{x}$ vectors according to their weights, and calculate $\boldsymbol{\Omega}$ using the training $\mathbf{x}$ vectors. Then, for each of the test set $\mathbf{x}$ vectors, we find the 10 nearest neighbors in the reduced dimension space. We predict the held out y values by taking a weighted average of the 10 nearest neighbor y values, weighted according to distance. We should note that in this case, the reduced dimension space $\mathbf{z}$ is produced using the eigenvectors of $\boldsymbol{\Omega}$ weighted by their corresponding eigenvalues, in order to make distance between $\mathbf{z}$ vectors more meaningful.

For each fold, we test a suite of τ and κ values in a stepwise fashion. We select a range of values on the $\log_{10}$ scale where all combinations of τ and κ are tested. Then, if there is a clear minimum holdout mean squared error (MSE) that is not an edge case, we select it as the minimum. If the minimum is on an edge, we explore further in that direction until the minimum is not an edge case.

After each fold is optimized, the optimal τ and κ values can be averaged across folds.

3.3 Interpretation

The $\mathbf{b}$ vector is interpretable when thought of as weights in the linear combination of $\mathbf{x}$ that results in an informative new variable (z), implying that large (absolute value) weights indicate more important parts of $\mathbf{x}$ (assuming $\mathbf{x}$ is standardized). This can be used as a sensitivity gage, where larger $\mathbf{b}$ values indicate a greater contribution of the corresponding $\mathbf{x}$ dimensions. Further, if $b_i > 0$ and $b_j < 0$, we can expect that increasing b_i while decreasing b_j (or vice versa) results in the same z value and hence the same y value. Smooth $\mathbf{b}$ vectors make sense when $\mathbf{x}$ is smooth, as we expect features of the $\mathbf{x}$ curves to influence the output rather than particular variables like x_i. In other words, if $\mathbf{x}$ is smooth we expect x_i and x_{i+1} to have similar effect on y. Without imposing smoothness on $\mathbf{b}$, the result is a great deal less interpretable as it may indicate x_i is important while x_{i+1} is not, or may have positive b_i and negative b_{i+1} in order to dampen the effect of highly correlated x_i and x_{i+1}. A ragged $\mathbf{b}$ may be equally good at reducing the dimension of $\mathbf{x}$, but is difficult to understand.

4 Incorporating Experimental Knowledge

A simple, but effective way to incorporate experimental knowledge into our simulator dimension reduction is to use importance sampling.

4.1 Importance Sampling

If we have a vector $\mathbf{x} \sim p(\mathbf{x})$, where $p(\mathbf{x})$ is difficult to sample, we can use importance sampling (Geweke (1989)) to weight independent samples from a similar distribution, $q(\mathbf{x})$, so that they are a weighted sample from $p(\mathbf{x})$. For instance, to find the expectation of a function $f(\mathbf{x})$,

$$\begin{aligned} E(f(\mathbf{x})) &= \int_{\mathcal{X}} f(\mathbf{x})p(\mathbf{x})d\mathbf{x} \\ &= \int_{\mathcal{X}} \frac{f(\mathbf{x})p(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x})d\mathbf{x} \\ &= E_q\left(\frac{f(\mathbf{x})p(\mathbf{x})}{q(\mathbf{x})}\right) \end{aligned}$$

which can be approximated with

$$E_q\left(\frac{f(\mathbf{x})p(\mathbf{x})}{q(\mathbf{x})}\right) \approx \frac{1}{M}\sum_{i=1}^{M}\frac{f(\mathbf{x}_i)p(\mathbf{x}_i)}{q(\mathbf{x}_i)} = \frac{1}{M}\sum_{i=1}^{M} f(\mathbf{x}_i)w_i$$

where $\mathbf{x}_1, \ldots, \mathbf{x}_M$ are independent samples from $q(\mathbf{x})$ and $w_i = p(\mathbf{x}_i)/q(\mathbf{x}_i)$ is called the importance weight. Now, when we are seeking a posterior expectation, we have that $p(\mathbf{x}) = \pi(\mathbf{x}|D) := L(D|\mathbf{x})\pi(\mathbf{x})/c$ where D denotes data we are conditioning upon, $L(\cdot)$ denotes the likelihood, $\pi(\mathbf{x})$ is the prior, and c is the normalizing constant. If we use $q(\mathbf{x}) = \pi(\mathbf{x})$, the prior, then the weights are

$$w_i = \frac{\pi(\mathbf{x}_i|D)}{\pi(\mathbf{x}_i)} = \frac{L(D|\mathbf{x}_i)\pi(\mathbf{x}_i)}{c\pi(\mathbf{x}_i)} = \frac{L(D|\mathbf{x}_i)}{c} \propto L(D|\mathbf{x}_i).$$

We use these weights, $w_1, \ldots, w_M$, to reweight our prior samples conditional on criticality experiments. For instance, simulations of Jezebel ought to produce k_{eff} near one with, perhaps, the probability of getting k_{eff} of 1 ± 0.0023 constant, and anything more extreme being less probable until reaching probability zero beyond 1 ± 0.009. This trapezoid-shaped distribution, called $h(\cdot)$ (shown in Figure 5), is then the likelihood. Hence, $w_i \propto L(D|\mathbf{x}_i) = h(f(\mathbf{x}_i))$, where f is the code that results in $y = f(\mathbf{x})$. We use the Jezebel simulations to obtain weights that we use to resample the simulations before performing dimension reduction. That way, the dimension reduction applies to the most pertinent part of the inputs.

Importance sampling falls apart when there are many small weights and a few large weights. The "effective sample size," estimated with $ESS := (\sum_{i=1}^{M} w_i^2)^{-1}$, is a gage of how many independent samples the weights effectively produce. In our case, weighting according to the Jezebel experiment trapezoid distribution above results in $ESS = 4453$, which is a sufficiently large sample to implement our dimension reduction techniques.

5 Comparison of Dimension Reduction Methods

For each of the nine GRSIR methods ($\mathbf{\Omega}_0$ through $\mathbf{\Omega}_8$), Figure 6 shows the distribution of holdout data RMSEs for the 10 cross-validation folds. This also shows the distribution of RMSEs for the active subspace approximation as well as for varying numbers of principal components. This plot demonstrates that the PCA approach is not well suited to this dataset. The dominant modes of variation in y do not correspond to variation in the first

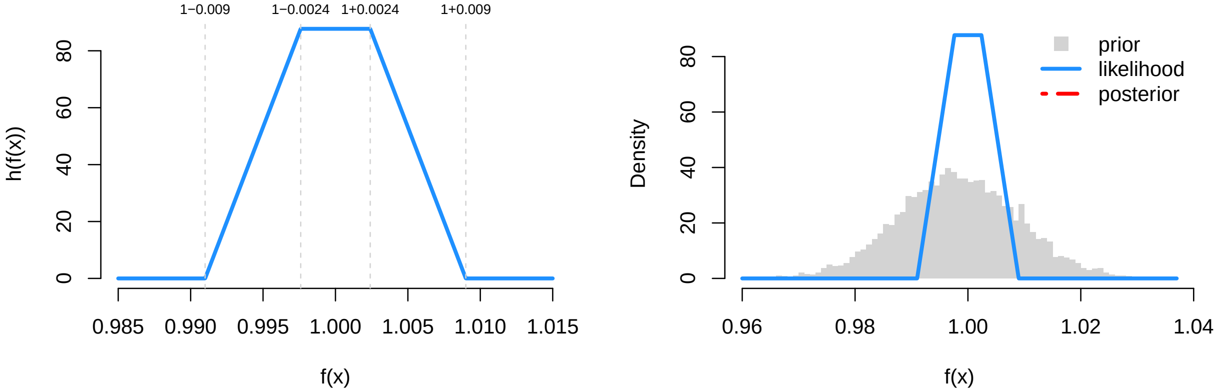

Figure 5: Left: the trapezoidal likelihood function $h(\cdot)$. Right: a comparison of the prior $f(\mathbf{x})$ values, where $\mathbf{x}$ is sampled from the Gaussian distribution described in Figures 1 and 2, and the posterior $f(\mathbf{x})$ values, which is the prior reweighted by the likelihood.

few principal components. Hence, all the GRSIR methods result in RMSEs more than an order of magnitude smaller than the PCA approach. Further, all the GRSIR methods only use one active dimension, where we show up to the first 10 principal components being used. A dimension reduction that results in $d = 1$ is much more efficient that one that results in $d = 10$. This plot also demonstrates that the active subspace approximation is well suited to this data, though it is not very different from the GRSIR approaches. Particularly, the GRSIR approaches that induce smoothing result in a simpler and smoother **b** vector, which we are more inclined to believe has physical meaning than more ragged **b** vectors. Hence, we use the smooth Tikhonov **b** vector for dimension reduction.

Figure 7 shows what the difference would be when choosing between using the first principal component for **b**, the smooth Tikhonov **b**, and the active subspace approximation for **b**. The right two plots show that the smooth Tikhonov and active subspace approximation methods result in an almost noiseless relationship between the corresponding z and y (i.e., the function g has little noise). The z-y relationship resulting from the first principal component has much more noise. This kind of noise is induced by the method of dimension reduction rather than the underlying high dimensionality of the data, and is thus avoidable. Hence, the first principal component provides a poor summary of the active dimensions.

Figure 8 shows the smooth Tikhonov **b** vector, which indicates the relative contribution of the corresponding cross sections, nubar, and PFNS. This indicates that the elastic, inelastic, and fission cross sections play dominant roles while the n2n, n3n, and capture cross section play little role. PFNS and nubar are moderately important. The low energies are unimportant for explaining the variation of the simulated k_{eff} values (conditioned on the Jezebel experiment). However, the very highest energies are also mostly unimportant.

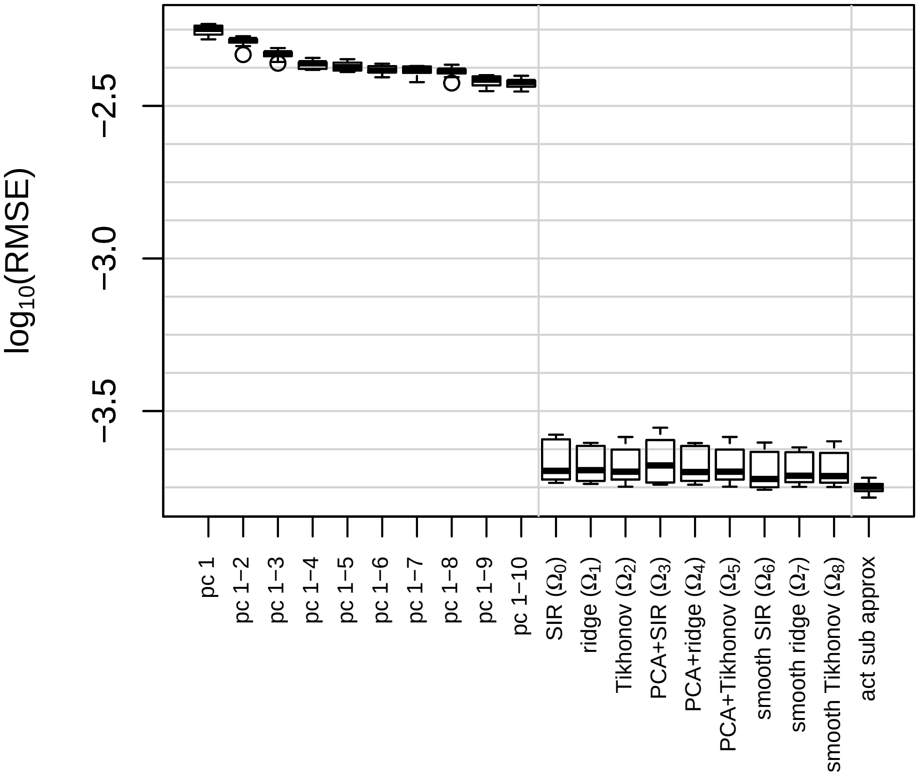

Figure 6: RMSEs for each dimension reduction method over 10 cross-validation folds, where the g function ($y = g(\mathbf{z})$) is estimated with nearest neighbors.

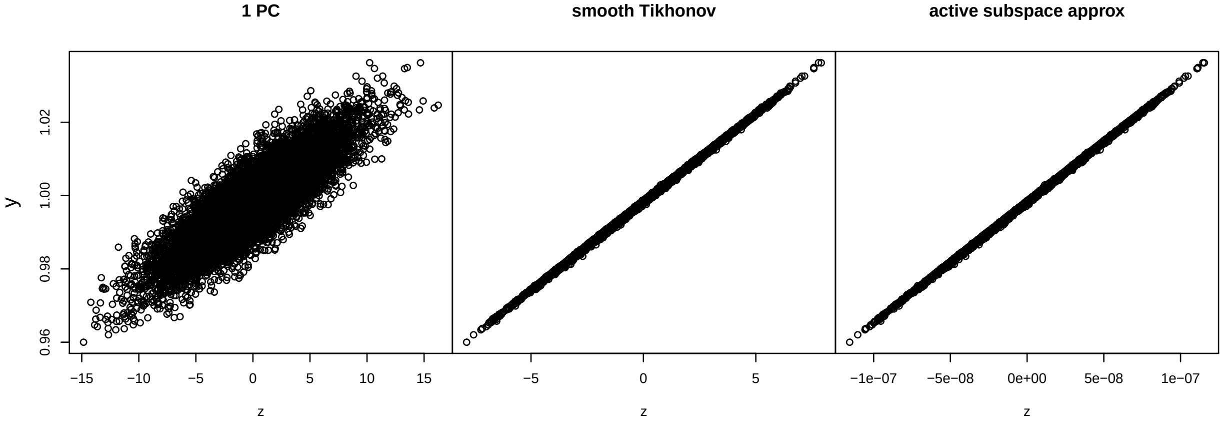

Figure 7: Plots of the z-y relationship (holdout data), where $z = \mathbf{b}'\mathbf{x}$ for a few different $\mathbf{b}$ vectors obtained using different dimension reduction methods.

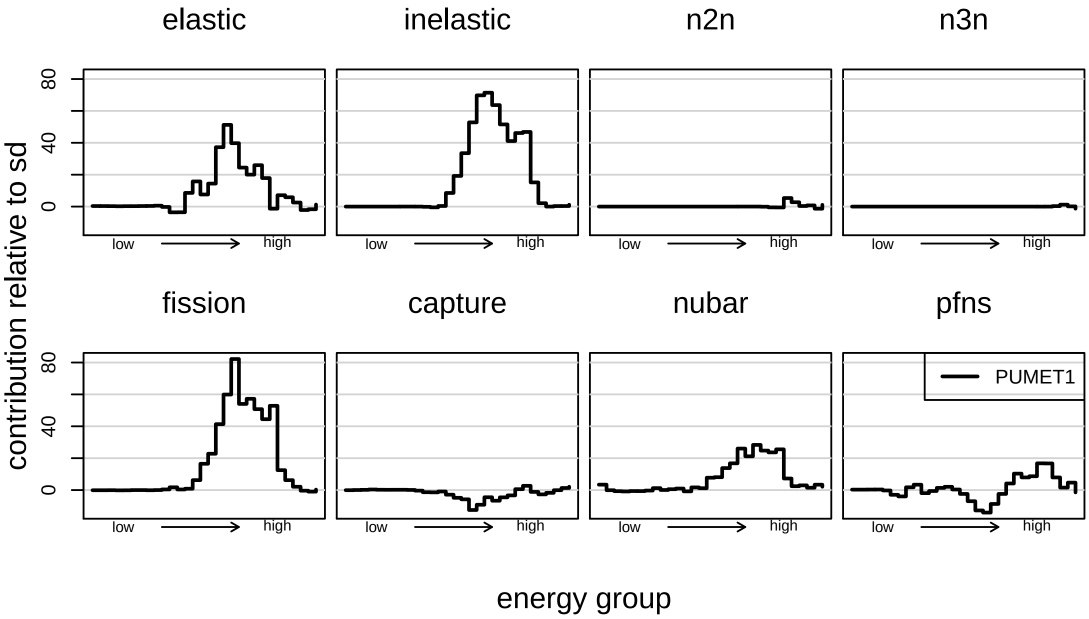

Figure 8: The smooth Tikhonov **b** vector, which indicates how cross sections are weighted to obtain the active variable.

6 References

Bernard-Michel, C., Gardes, L., and Girard, S. (2009), "Gaussian regularized sliced inverse regression," *Statistics and Computing*, 19, 85–98.

Constantine, P. G., Dow, E., and Wang, Q. (2014), "Active subspace methods in theory and practice: applications to kriging surfaces," *SIAM Journal on Scientific Computing*, 36, A1500–A1524.

Constantine, P. G., Eftekhari, A., and Wakin, M. B. (2015), "Computing active subspaces efficiently with gradient sketching," *arXiv preprint arXiv:1506.04190*.

Cook, R. D. (2007), "Fisher lecture: Dimension reduction in regression," *Statistical Science*, 1–26.

Donoho, D. L., et al. (2000), "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Math Challenges Lecture*, 1, 32.

Geweke, J. (1989), "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica: Journal of the Econometric Society*, 1317–1339.

Li, K.-C. (1991), "Sliced inverse regression for dimension reduction," *Journal of the American Statistical Association*, 86, 316–327.

Russi, T. M. (2010), "Uncertainty quantification with experimental data and complex system models," unpublished Ph.D. dissertation, UC Berkeley.